J. R. Statist. Soc. A,
(1973), **136**, *Part* 1, p. 21

Subjective Probability and its Measurement†

By J. M. Hampton, P. G. Moore and H. Thomas

London Graduate School of Business Studies

Summary

Subjective probability is defined and its place in decision analysis, with special reference to business problems, is identified. The literature on its measurement is critically analysed both for the single decision-maker and group. The use of direct fractile assessment and the Delphi Technique are felt to be some of the more tenable of the methods reviewed. An account of some of the behavioural aspects of decision-making with a résumé of the "risky" shift theories is included and their implications are discussed. Some practical guidelines and suggestions for further research are indicated.

Keywords: SUBJECTIVE PROBABILITY; BAYESIAN DECISION ANALYSIS; INDIVIDUAL DECISION-MAKER; PDF–CDF ASSESSMENTS; PSYCHOMETRIC RANKING; HFS; EPS; SCORING RULES; GROUP CONSENSUS; DELPHI TECHNIQUE; SEU MODEL; CONSERVATISM; PIP; CHOICE SHIFTS; SPAN

1. Introduction

Controversy exists about the meaning of probability in general, and subjective probability in particular, so that it is necessary to begin this survey by making clear our own interpretation of these words and its relation to other views held.

According to Laplace (see 1951 translation) probability is defined as the ratio of the number of cases favourable to the event to the number of all possible cases; each case is assumed to be equally likely. This definition is commonly referred to as the classical definition. Another approach is that of the "frequency" or "empirical" definition, commonly ascribed to Venn, whereby probability is defined as the limiting value of the relative frequency of the event concerned as the number of trials, in which the event concerned is one possible outcome, increases indefinitely.

On the other hand, Bruno de Finetti (1964) believes there is no need to assume that the probability of some event has a uniquely determinable value. His philosophical view of probability is that it expresses the feeling of an individual and cannot have meaning except in relation to him. De Finetti's detailed work has formed the basis of the subjective, or "personalistic" approach to probability, whereby subjective probability is considered a quantification of uncertainty. It represents the extent to which a coherent person believes a statement is true, based on the information available to him at that time. To be coherent he is required to make his assessments consistent with each other, such that fundamental contradictions do not exist among them. This coherence constraint ensures the applicability of the convention axioms of probability to the manipulation of these quantities. The postulates of coherence are such that it is impossible to set up a series of bets against a person obeying the postulates in such a manner that the person is sure to lose, regardless of the outcome of the events being wagered upon.

† A companion paper "Utility and its Measurement" is to appear in the next issue of the *Journal.*

The inability of the earlier, classical and frequentist, definitions to deal with statements referring to unique events indicates why the probability required in a decision analysis is necessarily one that has been subjectively assessed. The nature and likely occurrence of the events in the analysis will be within the sphere of experience of the decision-maker (d/m); his past knowledge is valuable and by use of his judgment his experience is explicitly incorporated into the analysis. These assessments are one of two subjective measurements required for decision analysis based on a Bayesian framework. An outline of the analytical approach (a fuller discussion is given, *inter alia*, by Raiffa, 1958) is as follows:

(a) list all possible decisions and events,
(b) by use of a decision tree identify the outcomes of each sequential decision-event combination,
(c) assess the worth (measured in utility terms) of these outcomes,
(d) assess the probability of the occurrence of the events, and finally,
(e) work back from the outcomes calculating expected worth, eliminating dominated paths, i.e. those with acts which have lower expected utility, to obtain the optimum initial strategy.

This paper surveys and analyses critically the literature on the measurement of subjective probabilities, for both the individual and group concensus situations. A review of relevant psychological studies on the actual behaviour of persons has been included, and suggestions are made for the direction in which further research will be conducted on the results of these prescriptive and normative approaches. Although the emphasis of the paper is for the consideration of business decision problems the concepts discussed are widely applicable.

2. Decision-maker: Individual

2.1. *Methods of Assessment*

Point estimation

In general there are two approaches to the measurement of the d/m's subjective probability for the occurrence of a single event. He may be directly interrogated concerning his judgments, or his assessments may be inferred from his preferences between possible decisions.

(a) The simplest approach is to ask the d/m what is his probability for an event. He may be asked to express his beliefs by stating a number between 0 and 1; or by a more visual response mode—for example dividing a line into two lengths corresponding to the relative probability of the two disjoint events.

(b) Savage (1954) argued in favour of indirect methods, in that measurement error, that is the discrepancy between reported and actually held judgments, will occur in such directly assessed probabilities and a more indirect method will obviate this disadvantage. A procedure called the gamble or lottery technique was used by him to obtain the subjective probability of the occurrence of single events. The d/m is required to answer a series of yes–no questions phrased in terms of simple betting odds. His subjective probability for the occurrence of the event is then implied from the odds required to make him indifferent between two offered bets. For example, the d/m is given an opportunity of accepting a gamble paying £50, say, if his sales next month exceed 1,000 units and nothing otherwise. Alternatively, he can accept £y ($y < 50$) without contingencies. Ideally there exists a value for y (for $y = Y$, say) which makes the two bets equally attractive to him. According to Bayes the expected value (EV) of the gamble is given by

$$EV = p50 + (1-p)\,0,$$

where p = probability that sales next month exceed 1,000 units. If the d/m is indifferent between receiving £Y for certain and the gamble, we have

$$Y = p50 + (1-p)\,0$$

or

$$p = Y/50, \text{ where } Y \text{ is known by enquiry.}$$

For such a method to be generally applicable there are some criteria which have to be satisfied: namely that there are only a few assessments to be made and that the d/m has both the patience and the time to work through the procedure. However, the main criticism of the method is the introduction of gambling terminology. Much of the work in the past in the field of measurement of an individual's subjective probability has used this approach. In agreement with current opinion, we believe that "special" gambling behaviour will be involved in betting situations which might mean that the quantity measured is not just the d/m's degree of belief about the occurrence of the event being considered, but might include his risk attitudes and propensity to gamble.

The concept of an "equivalent urn" (Raiffa, 1968) has been used to quantify a d/m's judgments concerning the chances of the occurrence of single events. The urn contains several coloured balls, each one is "labelled" with the name of one of the events (it may be simply E occurs, E does not occur; or "high sales", "medium sales", "low sales", etc.). The d/m has to assess what proportional mix of balls bearing the various labels will make the urn equivalent to his feelings concerning the likely occurrence of the events. This simple indirect method is effective for assessing a small number of events, but the task becomes impossible for a large number; it is then that the d/m has to consider assessing a distribution.

Assessing a distribution

In a business environment it will be extremely rare that the decision analysis will require the probabilities of a few events only at each fork. The fourth step in the analysis delineated in Section 1 as:

(d) assess the probability of the occurrence of the events, will often, in practice, require the assessment of many possible values of some uncertain quantity. The problem of assessing probabilities for a very large number of possible values of an uncertain quantity becomes one of assessing a distribution. The d/m must first decide how he wants his individual probabilities to be related to each other, after which he can assess just a few probabilities and then construct a curve to represent his judgements concerning the whole continuum of possible values.

The methods that have been suggested for such assessments are the use of probability and cumulative density functions, direct judgmental curve fitting, the smoothing of historical data and application of psychometric ranking and the use of hypothetical sample information (the HFS and EPS methods); these will be discussed in turn.

(a) *Probability and cumulative density functions*

For the direct fractile assessment of the CDF, Morrison (1967) designed the following questions to obtain the fractiles (F).

Qu. 1. At what value of the variable, $F(50)$, do you feel that there is a 50 per cent chance that the true value of the variable will be below $F(50)$?—to establish the value at which CDF = 0·5.

Qu. 2. Given that the true value of the variable is below $F(50)$ at what value of the variable $F(25)$ do you feel there is a 50 per cent chance that the true value of the variable will be below this value?—establishing the value for CDF = 0·25.

Qu. 3. Given that the true value is above $F(50)$ at what value of the variable $F(75)$ do you feel there is a 50 per cent chance that the value of the variable will be below this value?—establishing the value for CDF = 0·75.

From the set of discrete points obtained from these, and some additional questions on the same lines, an approximate curve for the CDF is drawn. An assessment of the PDF can be made following a similar method.

Schlaifer (1969) describes two methods:

(b) Direct judgmental curve fitting,
(c) Smoothing of historical data,

depending upon whether or not the d/m has any quantitative evidence available to him.

(b) *Judgmental curve fitting*

If the d/m believes that:

(i) the probabilities of the individual values of the uncertain quantity should fall off smoothly to either side of a single mode;
(ii) most of the probability should be concentrated within a fairly small interval around the mode;
(iii) there should be some probability fairly far out in the "tails";

then his judgments as to the occurrence of all the possible values of the variable, demand say, may be represented by a mass function of the following general shape shown in Fig. 1.

Such a method demands the assessments of very small probabilities for the definition of the complete curve, which constitutes a very difficult task for the d/m. Because of this difficulty Raiffa (1968) and Schlaifer (1969) believe that the d/m will have more intuitive feeling for assessing the probabilities for cumulative events.

If the d/m accepts the general shape of Fig. 1 to represent his mass function, the general shape of his cumulative function is implied from it. Namely, the slope of the cumulative curve must be zero at the origin, increase to a maximum and then decrease. Hence the d/m is required only to make assessments of the type: "I believe that there is a 75 per cent chance that the demand will be less than or equal to 1,000"—which are of a sort that he can make without undue difficulty because, except for the extreme ends, the variable d will have values that are neither too small nor too large to have a real intuitive feeling.

(c) *Smoothing of historical data*

The d/m is required to base his subjective assessments on the knowledge available to him at the time of decision; in the event of historical data being available he should use such evidence in his assessment of the probability distribution.

The approach is essentially the same whether the probability mass or cumulative function is assessed. The quality of the data will generally determine which form should be used; for example sparse data suggest the assessment of a cumulative function. The historical relative frequencies for every possible value of the unknown

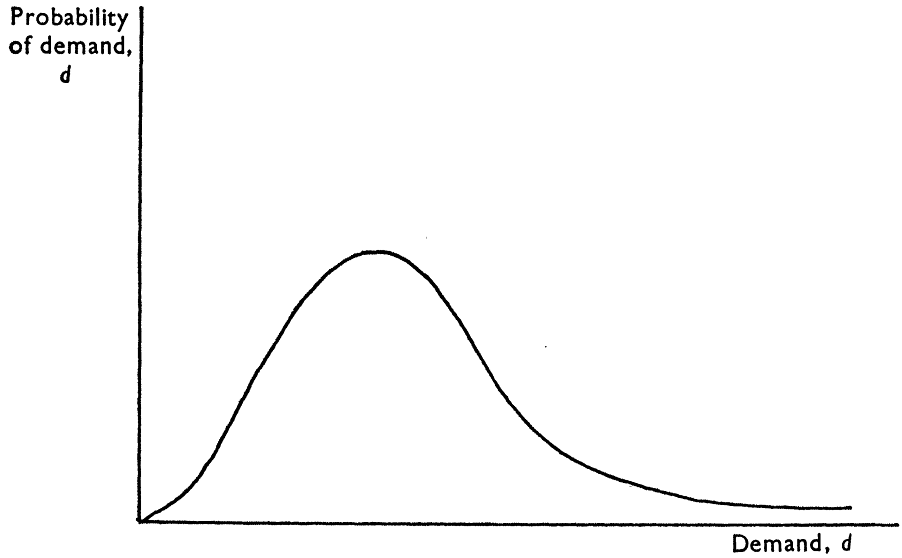

FIG. 1. General shape of probability distribution.

are plotted and a smooth curve whose general shape agrees with the d/m's general beliefs about the long-run behaviour of the process, but which remains as close to the individual points as possible, is drawn. The general shape will be unimodal and smooth if the d/m judges the values to be generated by a random process consisting of a "basic" and a "random" element, that is the effect of a large number of small independent factors. The d/m then uses this estimated long-run frequency distribution as a guide for his current decisions.

(d) *Psychometric ranking*

Smith (1967) proposed the use of a psychometric ranking technique for the assessment of a manager's beliefs relating to business decisions. We follow his approach of describing the method by an example. The manager is faced with the task of trying to assess the likely market share for a new product. The possible range of the market share is clearly from 0 to 100 per cent and is first broken up into 10 intervals. Smith maintains that, rather than trying to assess the relative probabilities of each interval directly, the manager should be asked to rank the various intervals in ascending order according to their expected relative probabilities of occurrence. Smith believes it feasible to require the assessor to repeat the procedure, this time using the first differences of these intervals.

Suppose that the manager believes the following rankings should be assigned to the various possible intervals:

Interval	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
Ranking	1	2	7	8	10	9	6	5	4	3

and that the differences should be ranked as follows:

Interval	0–10	10–20	90–100	80–90	70–80	60–70	20–30	30–40	50–60	40–50
Ranking	1	2	2	4	5	6	7	8	9	10
Ranking of differences	1	3	2	9	6	5	4	8	7	

(1)

Smith suggests the application of Kendall's (1962) quantification of rankings to translate the information imparted by the manager's rankings into relative probabilities. In order to quantify $(n+1)$ rankings Kendall recommends that the n first differences are assigned values which are in the following proportions:

$$\frac{1}{n}\left(\frac{1}{n}+\frac{1}{n-1}+\ldots+\frac{1}{2}+\frac{1}{1}\right)$$

$$\frac{1}{n}\left(\frac{1}{n}+\frac{1}{n-1}+\ldots+\frac{1}{2}\right)$$

$$\vdots$$

$$\frac{1}{n}\left(\frac{1}{n}+\frac{1}{n-1}\right)$$

$$\frac{1}{n}\cdot\frac{1}{n}.$$

In the example we have nine differences and hence:

$$\frac{1}{9}\left(\frac{1}{9}+\frac{1}{8}+\frac{1}{7}+\ldots+\frac{1}{2}+\frac{1}{1}\right) = 7{,}129/22{,}680$$

$$\vdots$$

$$\frac{1}{9}\left(\frac{1}{9}+\frac{1}{8}\right) = 595/22{,}680$$

$$\frac{1}{9}\frac{1}{9} = 280/22{,}680$$

giving the expected proportions as

$$280,\ 595,\ 953,\ 1{,}375,\ 1{,}879,\ 2{,}509,\ 3{,}349,\ 4{,}609,\ 7{,}129. \qquad (2)$$

The manager is then required to assess the "end" probabilities, that is the probabilities of the product capturing less than 10 per cent of the market and 40–50 per cent of the market. If he assesses these two probabilities to be 0·005 and 0·20 respectively, the range of probabilities is hence 0·20–0·005 = 0·195. The total range of the values is 22,680 so that the proportions in (2) are multiplied by 0·195/22,680 to obtain 0·0024, 0·0082, 0·0051, 0·0613, 0·0216, 0·0162, 0·0018, 0·0396, 0·0288 (re-arranged into the order as specified by (1)). Finally, these are accumulated to obtain 0·0050, 0·0074, 0·00156, 0·0207, 0·0820, 0·1036, 0·1198, 0·1136, 0·1712, as the relative probabilities of the product capturing that percentage of the market as designated by the intervals. Once normalized, these probabilities are plotted as a histogram which is smoothed into a frequency curve. Smith (1967) suggests the use of ranking techniques for the assessment of probabilities finding, as he does, such a method logical, definitive and consistent. This method was developed from successful applications in psychology of the concept of quantifying rankings (e.g. Coombs, 1964). Further work in this area has come from both Kruskal (1964) and Shepard (1966) in their discussions of non-metric multidimensional scaling. Although the application of ranking techniques may have a potential use in the assessment problem, the authors believe this method involving the ranking of differences will be psychologically and intuitively meaningless to the decision-maker, and concur with Morrisons's (1967) critique that the practice of ranking these paired comparisons is likely to tax the latter's patience.

(e) *Hypothetical future samples (Good's device of imaginary results)*

Once the subject has made a single assessment for the proportion under consideration, the HFS method examines the effect of additional knowledge of the distribution. The subject has to imagine what effect the additional knowledge of a given random sample would have on his original probability. For example, suppose he had assessed the probability that a student selected at random is a male as p_1 (Winkler, 1967b), then he may be asked questions of the form:

"Suppose he had taken a sample of 100 students and 60 were male. Now what is the probability that one additional student to be chosen at random is a male?"

(f) *Equivalent prior sample*

Similarly, this method considers the effect of sample information. In this indirect method the subject is requested to give values for r and n which would be equivalent to obtaining a sample where r, of n chosen at random, were males. The ratio r/n should be close to his original value of p_1 where the larger the sample size he chooses for n indicates his confidence in his original assessment.

The viability of some of these methods in practice is considered in the following section.

2.2. *Empirical Studies*

In practice, the assessor of probabilities is commonly not a statistician and is unfamiliar with the formal notions of probability, so that the methods to be used to obtain his subjective probability assessments have to be comprehensible to him. On the basis of this criterion the more elegant theoretical suggestions (for example, judgmental curve fittings) have to be discarded in favour of a number of more basic approaches, which we shall now consider.

(a) Perhaps the main contribution to the measurement problem in practical contexts has been by Winkler (1967a, b). After a pilot study, Winkler selected four main assessment techniques for the measurement of an unknown proportion. They were as follows:

(i) CDF *Cumulative Distribution Function*—the fractiles of the distribution are assessed and the CDF is graphed.

(ii) HFS *Hypothetical Future Samples*—the effect of sample evidence on the decision-maker's assessments is considered.

(iii) EPS *Equivalent Prior Sample Information*—prior judgments are expressed in the form of an equivalent prior sample.

(iv) PDF *Probability Density Function*—the points on the probability decision function are assessed by direct interrogation regarding relative density and relative areas from which the PDF is graphed.

Winkler reports that varying degrees of success were obtained from the use of direct ((i) and (iv)) methods on the one hand and indirect ((ii) and (iii)) methods on the other. To reduce these inconsistencies Winkler used a "feedback" session. The subject is presented with his assessments and, after a discussion, is asked to write down one "best" assessment; the mere realization of the existing differences resulted in their reduction. Winkler concluded that these differences between the distributions can be explained in terms of conservatism in the subjects. They appeared to have difficulty in making full use of all the information they had available to reduce their uncertainty, and so their assessments obtained by the direct methods were diffuse distributions. By the same rationale they allocated smaller weightings to the HFS than they should and higher sample sizes to the EPS, not realizing the implied "tightness" of the distributions obtained. The subjects ranked the indirect methods higher than the direct methods on a clarity scale, which aims to reflect (i) their understanding of the technique and (iii) how easy it was to use.

Winkler found that of the direct assessment methods, the use of PDF was more intuitively appealing to his subjects: where inconsistencies existed it was the CDF that was altered, indicating that they were less sure of working with this distribution. He concluded that it was feasible to use these methods for the assessment of probabilities. He indicated the limitations of his study in that the assessments were for estimating proportion only; however, the authors feel that the use of direct fractile assessment, in particular, may well be the approach to adopt for decision analysis in a business environment where the estimate may be of a quantity rather than a proportion.

(b) Another study that should be considered is Allen's (1968), as an illustration of a different approach to the special case of non-repeatable business decisions. His approach is based on Shackle's (1961) credibility and potential surprise concepts. Shackle believes it impossible to list all the project alternatives, which is a prerequisite for assessing their relative probability of occurrence. Instead he submits that the criteria of project selection should be credibilities: they do not have to add to one and hence can be calculated on the available basis of incomplete experience.

A credibility forecast diagram (Fig. 2) is derived for each variable of interest by means of the following reasoning:

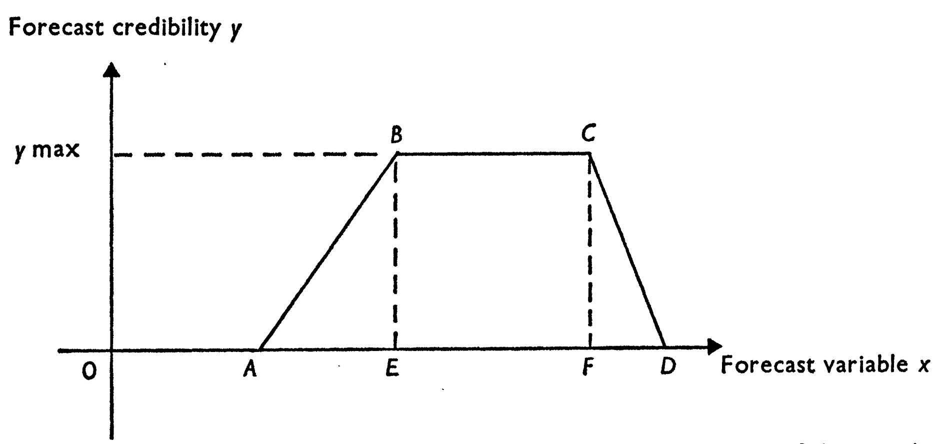

FIG. 2. Credibility forecast diagram. y max can be defined such that the area of the trapezium $= 1 = \frac{1}{2}(y \text{ max})\,(AD+BC)$ so that $y \text{ max} = 2/(AD+BC)$.

"On the basis of the information available to me at the present time I consider any value of x between B and C to be completely credible (or, alternatively, I would not be at all surprised if the actual value of x turned out to be anywhere between B and C). I consider it utterly incredible for x to have a value less than A or greater than D. . . ."

The potential surprise concept has been criticized for its ambiguity of meaning due to its qualitative nature. In addition the use of credibilities is incoherent based on the criteria mentioned earlier in this paper.

We note that the uncertainty described by equal credibility over a range is similarly expressed in Bayesian theory by the assignment of a diffuse prior. Such a distribution implies that every value outside this interval has a zero probability attached to it and is hence equivalent to Allen's credibility region: *EBCFE*.

2.3. *Scoring Rules*

An additional area of interest concerns the use of scoring rules to resolve difficulties in the assessment of the manager's own personal probabilities. These are measurement devices which encourage the assessor to report his personal assessments as accurately as possible, that is, to make the correspondence between the manager's assessments and judgments as close as possible.

A major difficulty in making assessments is the vagueness and ambiguity felt initially; in order to obtain assessments which are stable over time the manager needs an incentive which both increases his involvement in the task *and* encourages him to overcome the difficulty of obtaining the assessments from himself. Gauss (1821) was probably the originator of the view that economic losses or gains provide a sound basis for motivating experts to estimate accurately even in the most hypothetical cases. Some scoring rules incorporate this view and use money rewards while others rely on the score itself as an incentive. Whichever approach is adopted the scoring rule must encourage the manager to report his judgments accurately and the design such that the manager reveals his true judgments in his formal probability assessments (Toda, 1963; De Finetti, 1965).

Van Naerssen (1962) suggested a scale for the direct assessment of point estimates. He explains the method within the context of selection of candidates who may succeed or fail with relation to some specified criterion. With each candidate the selector has to choose between k categories of the subjective probability of success p. If he chooses p_j he is awarded x_j if the candidate succeeds and y_j if he does not. By assuming that:

(i) the selector will assign a probability of $1-p$ to the non-occurrence of an event if he assigns p to the occurrence of the same event, and
(ii) the utility of the score is a linear function of the score itself.

Van Naerssen derives quadratic solutions for the scale values x and y. He suggests scales like the following:

$100p =$	0–20	20–40	40–60	60–80	80–100
$x =$	-7	-3	0	2	3
$y =$	2	3	0	-3	-7

where the rationale is that, if the selector assigns a subjective probability for the success of a candidate as 0·5, he is not adding anything to an *a priori* assessment and hence earns no points.

Most of the current literature on scoring rules has come from the area of meteorology (see Epstein, Murphy, Stael von Holstein, Winkler and Murphy). One thing which becomes clear from studying this literature is that the size of the incentive in the rule varies according to the use to which you put the scoring rule. The difficulties encountered in the implementation of these rules as a training device include the possibility that the manager may not make careful assessments because the money reward is not sufficient to motivate him to take the assessment procedure seriously. In an actual business decision context, the cash incentives may be substantial because the analyst has an interest in making the fluctuations in wealth of the manager as large as possible to encourage him to consider seriously each assessment; however, the use of such large monetary incentives may mean that the manager's probabilities may reflect partly the non-linearity of his utility for money.

The aim, then, of such rules is to aid the assessor to make probability assessment; not to make him in any way get a more "correct" probability. The question of whether to use scoring rules or not becomes the problem of trading-off the advantages of possibly obtaining more consistent assessments against the disadvantages of additional complexity and the introduction of further conceptual situations. The choice of a scoring rule is a subtle and difficult task and practical research is needed to discover whether the use of scoring rules tends to evoke the correct amount of effort from the manager.

In conclusion the existing treatments of the measurement of an individual's subjective assessments in practical applications can be seen to be very limited. The concept is subtle and great operational difficulties are encountered in the application of proposed theoretical methods. We believe that it is essential to obtain consistent assessments and, of the approaches evaluated, that of direct fractile assessment seems to us to be the method most likely to achieve this end.

3. Decision-maker: Group

Formal Bayesian analysis envisages the identification of a single d/m's subjective probability assessment: in a business environment, however, this function will often be fulfilled by a group rather than a single individual. This raises a serious question as to the applicability of the analysis unless the individual assessments can be combined into a group assessment; or the members can be persuaded to modify their judgments so as to agree on a single assessment.

3.1. *Methods of Assessment*

When the individuals of the group each provide a separate subjective probability distribution representing their judgment, a consensus problem arises in aggregating these distributions into a single input for the analysis.

Winkler (1968) suggests a number of aggregation schemes based on a linear combination of the individual distributions. The weights used in such a linear combination reflect the rationale that some individuals might be regarded as better assessors than others. In the event that all are believed equally able, the weights are made equal. Winkler envisages a higher authority (D) to whom the group is responsible and who is required in one possible aggregation scheme to make the weight assessments.

Other weights proposed are those based on a self-rating ranking (by the members of the group) and on weights deduced from some comparisons of previously assessed distributions with actual outcomes. The scheme involving D's weightings can be recognized as a common informal system of committee decision where the chairman effectively weights opinions of the members in arriving at his assessment of the committee decision.

An alternative method, suggested by Winkler, is applicable if the judgments of the individuals are tantamount to sample evidence of the same kind as that to be analysed, then a natural-prior conjugate may be used to express the judgments. An intuitive explanation of a natural-conjugate prior is to say that the prior information can be thought of as being equivalent to sample information from the data-generating process of interest. Each individual's distribution will be a member of the natural–conjugate family of distributions and the distributions can be combined in a manner similar to successive applications of Bayes's theorem. These two different methods of aggregation produced varying results but Winkler maintains that D is responsible for the final assessment of the single distribution and he should use those methods which simplify his problem and appear relevant to him.

Some helpful comments regarding the consensus problem can be found in such references as Wilson (1968a) who is primarily concerned with the larger picture of the group decision problem. The essential structure is that a common decision α is chosen and depending on the outcome ϵ of the random variable $\tilde{\epsilon}$, the resulting payoff $x = p(\epsilon, \alpha)$ is shared, according to some rule, among the members. The paper gives the mathematical justification for the consideration of a group probability assessment and utility function. For each individual a "differential evaluation measure" is said to exist, which is a multiplicative function of his subjective probability distribution and marginal utility function. One interpretation of the existence of such a measure for the group is that it will be based on a surrogate marginal utility function for the group and a consensus, i.e. a group probability assessment. Wilson's treatment does not suggest methods for obtaining either of these desirable functions, but does explain the behavioural significance of their existence; namely that the group will be consistent (Savage's fourth postulate is thus satisfied).

3.2. *Empirical Studies*

The mathematical discussions of the group consensus problem assume that each member has already assessed their respective distributions. An alternative suggestion is to arrive at an agreement of opinion by encouraging each individual to reconsider his assessment, after presenting him with some feedback regarding the assessments of all the members of the group.

In his paper Winkler considers two such behavioural approaches: feedback and re-assessment (FR) and group re-assessment (GR). The former method is to present the member with the anonymous assessments of his colleagues together with their self-rated ranks and to obtain from him his individual reassessment. The method is repeated until a convergence of opinion is obtained. The GR approach is to have the individuals meet as a group to be presented with the feedback and for them to discuss the matter in order to arrive at a single consensus.

Dalkey (1969) criticizes such group discussions, believing them to be beset by the influence of dominant individuals, poor communication between members and by the distortion of individual judgments due to group pressure. Dalkey argues for the Delphi Technique which is a variant of Winkler's FR method. The chosen members of the group are "experts" on the subject under consideration. They independently develop their assessments, making explicit their underlying assumptions and sources of information and request any additional source which they feel would help refine their assessments. The feedback consists of the composite replies of their colleagues together with a list of underlying assumptions, and information requirements selected by the supervising experimenters. Only such theoretical assumptions that are considered to represent a consensus of a majority of respondents, and only such data as have been asked for by at least one respondent and is obtainable from reliable sources is included in the feedback. The identities of the members are withheld from each other to minimize personality interplay. Successive revisions of the original assessment are undertaken until a composite assessment can be compiled.

In Dalkey's (1969) study 10 experiments involving 14 groups of from 11 to 30 subjects were conducted and about 13,000 answers to some 350 questions obtained. The results indicated that more often that not the procedures made the group estimates more accurate and hence put the application of the Delphi Technique in situations where information is scarce on much firmer ground. Because the procedure lends itself to quantification Dalkey was able to obtain insights into the nature of the group information process. For example, the average error of the first revisions was found to be a linear function of the dispersion of the original assessments. A further result was that the average movement of opinion between the first and second revisions was a well-behaved function of the distance of the first revised answers from the group mean and the distance from the true answer. Further, Dalkey obtained meaningful estimates of the accuracy of the group response to a given question by combining individual self-ratings of competence on that question into a group rating; when combined with the two previously mentioned results of accuracy and standard deviation; the author of that paper maintains that accuracy scores could be attached to the results of the Delphi Technique.

It is felt (see the papers by Dalkey, Campbell and Hitchin, Dalkey and Helmer) that this technique reduces the influence of personalities by its anonymity, encourages better communication by the iteration of controlled feedback and minimizes the effects of group pressures by a statistical group response. Behavioural investigations into the effects of group interaction certainly lend support to Dalkey's criticisms of group discussions (see Section 4.2 in this paper) and we feel that the Delphi Technique is of considerable interest as a method for obtaining a group consensus using controlled feedback.

4. DESCRIPTIVE DECISION-MAKING: PSYCHOLOGICAL EVIDENCE

The decision-maker's assessments of probability are always assumed to be those of a rational idealized man. It is very natural, therefore, to include in this survey a section on the behaviour of real persons. The latter is seen to be of a very different nature when the decision-maker is represented by an individual or by an aggregate of individuals and the review is therefore under two sub-headings.

4.1. *The Individual*

The results from empirical studies carried out by psychologists that are felt to be relevant are those of: objective/subjective probability comparisons, factors that affect the assessments, the consistency of these assessments and finally the way in which people process information for the revision of their probabilities in the presence of new information.

As a validation of the use of subjective probabilities, the emphasis of many of the studies was on the comparison between the measured assessments and the "true" objective probabilities. These latter quantities were in fact the "usual" or conventional probabilities assigned to the occurrence of well-known events; for example, the rolling of dice. The studies all took place in a laboratory setting using events with such well-known probabilities, but once-only events have no conventional probabilities and hence in a business environment there is no basis for such comparisons.

It was found that subjects "overestimated" low probabilities and "underestimated" large ones. These results were obtained by Preston and Baratta (1948) who used an auction technique to obtain the indifference bids of their subjects, whose subjective assessments were inferred from these bids. Among the many later studies which confirmed the effect are those of Griffiths (1949), Vail (1952), Attneave (1953), and Sprowls (1953).

Tversky (1967) in an attempt to test utility theory discussed this observation. His experiment is representative of empirical studies on the conjoint measurement of subjectively assessed utility and probability. (See also Davidson *et al.*; Coombes and Beardslee.) In his experiment he obtained the subject's minimal price $M(x, f)$ for a gamble; that is the smallest amount of money for which a subject would sell his right to play in a gamble where he stood to win £x if f occurred and nothing if it did not.

Although Tversky chose to draw comparative conclusions between "objective" and "subjective" probability from his results, the real worth of his experiment was that the data suggested the multiplicative relation between utility and subjective probability embodied in the subjective expected utility principle. The calculation of a d/m's subjective expected utility (SEU) for a set of n independent and exhaustive outcomes x_j ($j = 1, \ldots, n$), is expressed by

$$\text{SEU} = \sum_{j=1}^{n} p_s(x_j)\, u(x_j),$$

where $p_s(x_j)$ = subjectively assessed probability of occurrence for the outcome x_j. Coombs (1964) believes these experiments support the use of an SEU model to account adequately for most of the data in simple gambling experiments. Its descriptive validity remains untested for real world problems due to the large number of unknown parameters and appropriate controls.

The early interest expressed by psychologists for the relationship between "objective" and "subjective" probability is of limited interest today if one accepts that decision analysis requires the assessment of a probability that reflects the beliefs of the decision-maker based on the information available to him at the time of decision. The concept of an underlying "true" probability is hence of little practical or theoretical interest. The relative frequentist definition of probability, however, necessitates that the introduction of any subjective element can only be considered as an estimate of the "true" objective probability. The formulation of decision analysis within a Bayesian framework means that the beliefs of the decision-maker are all that is relevant and these are what have to be assessed, they can then, in no sense, be estimates.

The relevance of the psychologist's analysis of the way in which people assess subjective probabilities is felt to be in relation to training, in trying to measure subjective assessments we have at least to be aware of some of the influence of human beliefs.

The previously cited article by Vail raises several questions. The degree of realism and revocability of the decision may mean that one form of subjective probability calculi will be invoked in a laboratory setting whereas a completely different form used in the boardroom. Vail also discusses the way a subject's perception of mathematical probabilities may be a function of the wording of the problem, time allocated for completion, etc., as cognisance that such faculty will be affected by haste, inability to perform complicated computations and the forgetting of essential data. Further factors considered to influence probability assessment include level of difficulty, information available and the amount of interaction of the problems (Howell, 1967). The amount of experience the decision-maker has both with respect to assessing probabilities and within his general business framework is believed by some to be of significance. Cohen and Hansel (1958) submit that age, experience and the number and value of the alternatives offered are factors of importance. Later studies (Strub, 1969) support the use of trained personnel in complex decision tasks. They apparently processed information better, were more aware of the meaning of the prior probability values and adopted a maximization principle more consistently in that they tried to obtain the optimum return from each decision task. The worth of business experience as an aid to making subjective assessments within the field has been disputed as an advantage: Martin (1969) believes a negative transfer of training occurred with his subjects where their experience had not aided them in their decision tasks. The "experienced" subjects were observed to process information poorly and endorse fewer correct hypotheses than the "naive" subjects (Martin's terminology).

However, we feel that the amount of knowledge the decision-maker brings to the assessment in the form of business experience is an important positive factor to the task: the difficulty experienced in the translation of such experience into subjective probabilities is felt to be the possible cause of the sometimes observed negative transfer of training. A study session on the concept of probability and the requirements for consistency would equip the decision-maker with the necessary prerequisites for a successful attempt at subjective assessments.

Training procedures and measurement techniques that lead to highly consistent subjective probability measures become a major objective in the light of past studies that have tested this assumption. Cohen *et al.* (1956, 1957) tested different groups of children and found that for less than five trials the probabilities did not consistently add to one. The theoretical justification for calling these quantifications of belief, probabilities, depends seriously on the truth of this assumption. Although a non-additive probability theory has been suggested (Edwards, 1962) which would not necessitate this constraint, this is felt to pose a more complex problem than that of securing consistent subjective assessments.

Psychologists have used Bayes's theorem as an optimum rule for processing new information to revise probabilities and have studied how human behaviour is seen to compare with this rule. A reluctance to extract as much certainty from the data as was theoretically implied by the Bayes theorem: called conservatism, has been demonstrated conclusively (as representative material see Edwards and Phillips, 1966; Kozielecki, 1970). In technical tasks where the number of alternatives become very large the contrary effect of radicalism was exhibited (Kozielecki, 1970). The explanations of the behaviour have been that either human beings cannot attribute the correct diagnostic value to the data (Beach, 1966; Lathrop, 1970) or that they evaluated the diagnostic value of the data correctly but failed to combine the probabilities successfully (Phillips, 1965; Edwards, 1966; Peterson and Phillips, 1966; DuCharme and Peterson, 1968).

If the use of Bayes's Theorem as a method of combining probabilities to produce a revised probability in the light of new information is accepted, the wealth of psychological literature on the sub-optimality of human information processing serves only to give weight to the opinion that this function should be removed from those expected to be performed by the decision-maker. A natural extension to this logic is to ask which functions in decision analysis are best carried out by the decision-makers themselves and which by the formal application of analysis techniques?

One suggestion for a man–machine interaction scheme is called Probabilistic Information Processing systems. The human element is to be present in the initial stages of the analysis to furnish the lists of possible hypotheses and acts, as well as to assess subjective likelihood ratios for each datum and each pair of hypotheses under consideration. These data are then used as input for a computer program.

Bayes's theorem is used to aggregate the various likelihood ratios into a posterior distribution that reflects the revision of beliefs about the hypotheses due to the consideration of all the available information. Other systems were compared with a PIP system in an extensive experimental study conducted by Edwards *et al.* (1968). The criterion for comparison was to assert that:

> "... the diagnostic system performs best which, in situations in which all systems end up in agreement about the right diagnosis, reaches that diagnosis on the basis of least information ...".

The conclusion of the paper was that PIP processed the information dramatically more efficiently than did any of the other methods considered. Apart from its interest as a method of information processing a by-product of the work on PIP is that feedback to humans on the current state of opinion of the system was seen to be detrimental to their performance.

In the light of this research it would appear that a combination of human judgment and Bayes's theorem processes information more effectively than either does alone. However, the distinguishing feature of most business applications is the absence of information. For example, in a decision concerning the launching of a new product the decision-maker may well have launched a new product before, and such experience will assist in making his prior assessment. But he is most unlikely to have on-going data, such as market reactions to the new product or the market's buying behaviour, so that the lack of data raises the difficulty in the decision problem. It can be seen that, in the event of information being available, the use of such systems as PIP certainly seem advisable but with special reference to business applications are of somewhat academic interest only.

These studies of an individual's behaviour lead us to the following conclusions. The failure of subjectively assessed probability to equal the "objective" probability of comparable events is not a disturbing result for these assessments are taken as a decision-maker's beliefs based on his present knowledge and hence there can in no sense be a "correct" probability. The discussion of factors that may influence assessments is felt to be of interest in two fields. Firstly, in the training of businessmen believed to be necessary to help them understand the calculus of probability and secondly, in the design of future experiments on probability assessments. For example, because of the influence of the degree of reality of the situation on the way in which the subjects make their assessments, further research into experimental techniques will need to be in real life decision problems rather than in the laboratory. The observed inconsistency of these judgments is a serious result for their use in probability analysis, however, and as has already been discussed, the use of better training and measurement techniques should secure consistent assessments. The explanation of the sub-optimality of human beings as information processors, because of their difficulty in aggregating probabilities, suggests the operational procedure of decomposing all compound events into components, subjectively assessing these and then using Bayes's theorem to combine them with information about the event in the form of likelihood ratios to produce the revised probability of complex events.

4.2. *The Group*

The effects of group interaction on decision-making have had to be seriously considered when the analysis is applied to a business context because of the frequency with which the "decision-maker" is in practice an aggregate of individuals. Social psychological research has provided explanations for the difference which is seen to exist between the individual recommendations and the group opinion (for a recent review of the topic see *J. Pers. & Soc. Psychol.*, 1971) and to provide methods for combining the former into the group consensus.

After a group discussion a shift in risk preferences compared to those initially held by the individual members of the group has been observed. The first researchers in this area (Stoner, 1961; Wallach *et al.*, 1962) employed a set of choice-dilemma questions to investigate the riskiness exhibited by groups after discussion compared to that shown by the individual members. A risk shift was found to have occurred taking the problem set as a whole. Later evidence (Brown, 1965; Pruitt & Teger, 1967), however, showing that shifts occur towards caution after group discussion of certain issues, suggests that the more general phenomenon of choice shift exists. Several theories have been developed to account for this choice shift; they may be categorised as:

(a) The diffusion of responsibility theory.
(b) The familiarization theory.
(c) The leadership theory.
(d) Value theories:
  (i) Social-comparison theory;
  (ii) Pluralistic-ignorance theory;
  (iii) Release theory;
  (iv) Relevant-arguments theory.
(e) Decision theories.

(a) The work of Wallach *et al.* (1962) explained the shift towards risk only. They suggested that group experience reduces anxiety about the possible negative consequences of making the risky decision because the responsibility is psychologically shared amongst the other members of the group. This reduction of anxiety makes it possible to accept the risky alternative at a lower probability of success. This "diffusion of responsibility theory" is held to be a very unlikely explanation.

Two additional early theories, namely the familiarization theory (Bateson, 1966) and the leadership theory (Marquis, 1962; Collins and Guetzkow, 1964) have also had little experimental support.

(b) The former maintains that increased familiarity with the decision problem makes people willing to take increased risks because of a general reduction in the uncertainty associated with it.

(c) The latter theory suggests that higher risk-takers are more persuasive in group discussions.

Both these theories have been conclusively rejected by the results of later experiments (Bateson, 1966; Lamm, 1967; Teger and Pruitt, 1967).

(d) The theories that attribute choice shifts to the operation of widely held human values have had the most support from later studies. (Pruitt and Teger, 1967; Stoner, 1968; Clark and Willems, 1969; Fraser *et al.*, 1970).

(i) Of these "value theories" the "social-comparison" version (Brown, 1965; Stoner, 1968; Jellison and Riskind, 1970) is perhaps the most doubtful. The individual is believed to appraise the positions of the other members and to choose his initial level of risk to be at or above what he believes to be the group average. If, in discussion, he finds himself below the group mean he will revise his decision in a risky direction. The evidence of later studies supports the following three other versions of the value theory.

(ii) The "pluralistic-ignorance theory" maintains that members of a group embrace one attitude to risk but believe their colleagues embrace another. An individual will then compromise between his ideal preference level and an assumed group standard. If discussion reveals that the other members are close to the individual's ideal, his revised decision will be closer to it.

(iii) Release theory describes riskiness as an ego-ideal and the attempts to take riskier preferences is interpreted as an effort to achieve self-esteem. The "Walter Mitty" effect is embraced with this explanation: where the d/m sees himself as a high risk-taker but acts cautiously when faced with an actual decision. When a high risk-taker is observed in the group he gives his colleagues the courage of their convictions; they are "released" from the social constraint of moderateness.

(iv) A fourth form of value theory, "relevant-arguments theory," holds that the dominant value or values in a decision problem elicit persuasive arguments in a group discussion that convince members to move further in the direction of these values. So that if the uncertainty in the decision problem propagates risky arguments in discussion a risky shift will be observed. The experiments of Myers and Bishop (1971) and Silverthorne (1971) yield impressive support to this theory.

(e) A few authors have employed utility based decision theory to explain shifts in risk preference. There are two forms:

(i) After a group discussion the individuals change the utilities that they had originally assigned to the outcomes of the available options, and

(ii) After a group discussion the individuals' assignments of utilities for the outcomes converge (Myers, 1970).

There is little evidence for or against these hypotheses but the work of Vinokur (1971) may be interpreted (Pruitt, 1971) as the amalgamation of hypothesis (i) and the "relevant-arguments" value theory approach. He concludes that arguments heard in a group discussion produce utility changes which in turn produce choice shifts. The "relevant-arguments theory" augments this explanation by specifying the precise impact of those arguments on the d/m.

Some general conclusions can be drawn from this work. Firstly, that contrary to early belief the mechanism underlying shifts on non-risk involving and risky items is the same. Secondly, that values underlie these shifts and lastly a two-process approach may well be necessary to explain the phenomenon.

Notwithstanding the presence of this tendency, the need for group consensus has led to the development of a theory for collective decisions. The contribution from the social/psychological science has been limited as might be expected from an essentially descriptive science. However, Coleman (1966) prescribed some ideas for a general theory. Essentially he applied the "perfect market system" of economics to the situation. Individuals were to exchange their partial control in some choices over issues which do not interest them for more weight on issues that do. A more procedural technique, called SPAN, for aggregating individual judgments into a group decision has been developed by MacKinnon (1966). SPAN is a mnemonic device for Successive Proportionate Additive Numeration. Each member of the group has 100 points which he can allocate to other members of the group or among the options under consideration. This cycle identifies each member's allocation pattern. Any points received by a member from the initial allocation of his colleagues are allocated among representatives and options in the proportions indicated by his initial allocation cycle. Iteration of the procedure is continued until all points "converge" among the options thus indicating their weightings and hence the final decision.

The rationale of the method is that it allows adaptive involvement of many beliefs and attitudes, perhaps only indirectly relevant to the decision-making, that the single vote rules out. If a member believes that one of his colleagues is better suited than himself to vote on an issue, he may allocate some of his voting power to the colleague he has selected as his representative; in the other event he may vote for the option or options he believes best. MacKinnon believes that the method will lead to a "vertical" progression of points to members of the group with the greater merit. However, one suspects that it may lead to the re-allocation of voting power based on the wrong criteria and furthermore these criteria are never stated explicitly. For example, a more timid member may be influenced by the confidence and degree of vocalization exercised in the past by a colleague on one of the options under consideration, and thus believe him to be a sound representative. In fact these criteria are in no way a measure of the representative's knowledge. The anonymity of the method, however, will minimize such action.

From this survey of group behaviour it appears that the group opinion will differ from those expressed by its constituent members because of processes which take place during discussion, that lead individuals to revise their original recommendations. The implications for decision theory in a business environment are complex. The early result that a risky shift always occurred from group discussion implied that groups should be used when risk taking was desired and individuals when caution was desired. In the light of later studies this naive conclusion is no longer valid. It now appears there are many dimensions upon which groups differ from the average of the individuals who comprise them, whilst social psychological research has yet to develop an efficient theory to account for the differences. The conclusion that utility changes accompany choice shifts emphasizes the need for an explicit utility function for the firm. In the absence of further evidence, the presence of a shift does not negate the use of aggregation schemes to combine the individual's recommendations into a group consensus. One such method SPAN has been considered but we feel that in the business environment the Delphi Technique already mentioned will be of more value.

5. Practical Guidelines

In company with a number of authors (Phillips, 1966; Winkler, 1967; Souder, 1969) who have experienced the difficulties encountered in measuring subjective probabilities, we feel that training plays a substantial role in the successful application of any of the methods considered.

The misconceptions held by managers on the meaning of probability and the way in which it can be used, is believed to be a major hurdle in obtaining meaningful assessments. A discussion of probability and its calculus can be presented as the language of probability that the manager needs to translate his relevant beliefs and available information into probability assessments.

It appears that training should develop the assessor's ability to decompose a problem into separate factors, so reducing the total assessment problem to the assessment of distributions for each of the component factors. The evidence of the literature suggests that large influxes of information do not, *per se*, aid the decision-maker's tasks; the necessity of being able to combine information and beliefs into a complex whole requires the understanding of probability fundamentals.

The need for the assessor to consider his feelings on the occurrence of some event independently of any desired outcome must be explicitly stressed, and the d/m made aware of the "optimism" of "wishful thinking" (Slovic, 1964) and of the "pessimism" of allowing the seriousness of the outcome bias his assessment (Croston and Gregory, 1969).

In connection with training devices an additional area of interest concerns the use of scoring rules to resolve difficulties in the assessment of the manager's own probabilities.

The frequent use of scoring rules is believed to aid improvement in the long-run consistency of probability assessments rather than to be methods, in themselves, for obtaining probability assessments. The authors feel that they may well have a place within a training session as described above.

6. CONCLUSION

The aim of this paper has been to summarize and assess the literature on the measurement of subjective probability for the application of decision analysis.

The nature of the uncertainty in business decision problems generally requires the assessments of distributions rather than single events. If the decision-maker can be identified as a single person the direct fractile approach with the PDF appears to be the most useful method. However, when a group fulfils this role the use of an aggregation scheme or of controlled feedback, such as the Delphi Technique, seems more appropriate.

The experience a decision-maker has in relation to the decision problem, its relevance to him, its degree of revocability and the amount of relevant information available to him are significant behavioural factors in his decision process. Training, including the use of scoring rules, and practice in the technique of assessing probabilities may well reduce inconsistencies in the decision-maker's assessments. Understanding the factors that operate during a group discussion will aid the choice of suitable methods for obtaining a decision from a group of people, while substantiation of Vinokur's "changing utility argument" indicates further research in the allied area of a group utility function.

The theoretical studies that have been completed suggest several avenues for investigation; however, the authors feel that only experience with relevant business problems and applications will indicate the most fruitful direction for further research.

REFERENCES

ADELSON, R. M. and NORMAN, J. M. (1969). O.R. and decision making. *Operat. Res. Quart.*, **20** (4), 399–413.

ALLEN, D. H. (1968). Credibility forecasts and their application to the economic assessment of novel research. *Operat. Res. Quart.*, **19**, Part 1, 25–42.
ATTNEAVE, F. (1953). Psychological problems as a function of experienced frequency. *J. Exper. Psychol.*, **46**, 81–86.
BATESON, N. (1966). Familiarisation, group discussion and risk-taking. *J. Exper. Soc. Psychol.*, **2**, 119–129.
BEACH, L. (1966). Accuracy and consistency in the revision of subjective probabilities. *IEEE Trans. Human Factors in Electronics*, H.F.E.–7, 29–37.
BROWN, R. (1965). *Social Psychology*. New York: Free Press of Glencoe.
CAMPBELL, R. M. and HITCHIN, D. (1968). The Delphi Technique: implementation in the corporate environment. *Man. Ser.*, Nov./Dec., 37–42.
CLARK, R. D., III and WILLEMS, E. P. (1969). Risk preferences as related to judged consequences of failure. *Psychol. Reps.*, **25**, 827–830.
COHEN, J., DEARNALEY, E. J. and HANSEL, C. E. M. (1956). The addition of subjective probabilities. *Acta Psychol.*, **12**, 371–380.
—— (1957). Measures of subjective probability. *Brit. J. Psychol.*, **48**, 271–275.
COHEN, J. and HANSEL, C. E. M. (1958). Subjective probability, gambling and intelligence. *Nature, Lond.*, **181**, 1160–1161.
COLEMAN, J. S. (1966). Foundations for a theory of collective decisions. *Amer. J. Soc.*, **71**(6), 615–627.
COLLINS, B. E. and GUETZKOW, H. (1964). *A Social Psychology and Group Processes for Decision-making*. New York: Wiley.
COOMBS, C. H. (1964). *A Theory of Data*. New York: Wiley.
COOMBS, C. H. and BEARDSLEE, D. C. (1954). Decision making under uncertainty. In *Decision Processes* (R. M. Thrall, C. H. Coombs and R. L. Davies, eds.). New York: Wiley.
CROSTON, J. D. and GREGORY, G. (1969). Critique of O.R. and decision-making. *Operat. Res. Quart.* (*U.K.*), **20**, 4, 415–420.
DALKEY, N. (1969). An experimental study of group opinion. *Futures*, September, **1**, No. 3, 408–426.
DALKEY, N. and HELMER, O. (1962). An experimental application of the Delphi method to the use of experts. *Man. Sci.*, **9**, 458–467.
DAVIDSON, D., SUPPES, P. and SIEGEL, S. (1957). *Decision-making: An Experimental Approach*, Ch. 2, pp. 19–30; 49–81. Stamford University Press.
DE FINETTI, B. (1964). Foresight: its logical laws, its subjective sources. In *Studies in Subjective Probability* (H. E. Kyburg and H. E. Smokler, eds.). New York: Wiley.
—— (1965). Methods for discriminating levels of partial knowledge governing a test item. *Brit. J. Math. Statist. Psychol.*, **18**, 87–123.
DION, K. L., MILLER, N. and MAGNAN, M. A. (1970). Cohesiveness and social responsibility as determinants of group risk taking. *Proc. Ann. Convn of Amer. Psychol. Ass.*, **5**, 335–336.
DUCHARME, W. M. and PETERSON, C. R. (1968). Intuitive inference about normally distributed populations. *J. Exper. Psychol.*, **2**, Part 1, 269–275.
EDWARDS, W. (1954). The theory of decision-making. *Psychol. Bull.*, **51**, No. 4, 380–417.
—— (1962). *Subjective Probabilities Inferred from Decisions*, pp. 93–158. Psychol. Rev., New York: Wiley.
—— (1966). Non-conservative probabilistic information processing systems final report. ESD Final Report, No. 058, pp. 93–22–F. Unpublished manuscript.
—— (1968). Decision-making—psychological aspects. *Int. Ency. Soc. Sci.*, **4**, 34–42.
EDWARDS, W. and PHILLIPS, L. D. (1962). Man as transducer for probabilities in Bayesian command and control systems. In *Human Judgments and Optimality* (M. W. Shelley and G. L. Bryan, eds.), New York: Wiley.
—— (1966). Conservatism in a simple probability inference task. *J. Exper. Psychol.*, **72**, No. 3, 346–354.
EDWARDS, W., PHILLIPS, L. D., HAYS, W. L. and GOODMAN, B. C. (1968). Probabilistic information processing systems: design and evaluation. *IEEE Trans. Systems Science and Cybernetics*, SCC–**4**, No. 3, 248–265.
EPSTEIN, E. S. (1969). A scoring system for probability forecasts of ranked categories, *J. Appl. Meteor.*, **8**, 985–987.
FRASER, C., GOUGE, C. and BILLIG, M. (1970). Risky shifts, cautious shifts and group polarisation. *Eur. J. Soc. Psychol.*, **1**, 7–30.

GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* Cambridge: M.I.T. Press.

GRAYSON, C. J. (1960). *Decisions Under Uncertainty: Drilling Decisions by Oil and Gas Operators.* Boston: Harvard University.

GRIFFITH, R. M. (1949). Odds adjustments by American horse-race betters. *Amer. J. Psychol.*, **62**, 290–294.

HOWARD, R. A. (1966). Decision analysis: applied decision theory. Proceedings of the I.F.O.R.S. Conference, Boston.

HOWELL, W. C. (1967). An evaluation of subjective probability in a visual decision task. *J. Exper. Psychol.*, **75**(4), 479–486.

JELLISON, J. M. and RISKIND, J. (1970). A social comparison of abilities interpretation of risk-taking behaviour. *J. Pers. & Soc. Psychol.*, **15**, 375–390.

KENDALL, M. G. (1962). Ranks and measures. *Biometrika*, **49**, 133–137.

KOZIELECKI, J. (1970). Psychological characteristics of probabilistic inference. *Acta Psychol.*, **34**, 480–488.

KRUSKAL, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, **29**, 115–130.

LAMM, H. (1967). Will an observer advise high risk-taking after hearing a discussion of the decision problem? *J. Pers. & Soc. Psychol.*, **6**, 467–471.

LAPLACE, P. S. (1951). *A Philosophical Essay on Probabilities.* Translated by F. W. Truscott and F. L. Emory. New York: Dover.

LATHROP, R. G. (1970). Response acquisition in multi-dimensional riskless tasks. *Proc. Ann. Conv. Amer. Psychol. Ass.*, **5**, 19–20.

LEE, W. (1971). *Decision Theory and Human Behaviour.* New York: Wiley.

LINDLEY, D. V. (1971). *Making Decisions.* New York: Wiley, Interscience.

MACKINNON, W. J. (1966). Development of the SPAN technique for making decisions in human groups. *Amer. Beh. Sci.*, **9**(9), 9–13.

MARQUIS, D. G. (1962). Individual responsibility and group decisions involving risk. *Indus. Man. Rev.*, **3**, 8–23.

MARQUIS, D. G. and REITZ, J. H. (1969). Effect of uncertainty on risk-taking in individual and group decisions. *Beh. Sci.*, **14**(4), 281–288.

MARTIN, D. W. (1969). Data conflict in multinomial decision tasks. *J. Exper. Psychol.*, **82**, 4–8.

MILLER, H. (1970). Is the risky shift the result of a rational group decision? *Proc. Ann. Con. Amer. Psychol. Ass.*, **5**, 333–334.

MORRISON, D. G. (1967). Critique of ranking procedures and subjective probability distributions, *Man. Sci.*, **14**, 253–254.

MYERS, D. G. and BISHOP, G. D. (1971). Enhancement of dominant attitudes in group discussion, *J. Pers. & Soc. Psychol.*, **20**, No. 3, 386–391.

PETERSON, C. R. and PHILLIPS, L. D. (1966). Revision of continuous subjective probability distributions. *IEEE Trans. Human Fac. in Elec.*, HFE–7, 19–22.

PHILLIPS, L. D. (1965). Some components of probabilistic inference. Ph.D. Dissertation, University of Michigan.

PRESTON, M. G. and BARRATTA, P. (1948). An experimental study of the auction value of an uncertain outcome. *Amer. J. Psychol.*, **61**, 183–193.

PRUITT, D. G. (1971). Toward an understanding of choice shift in group discussion. *J. Pers. & Soc. Psychol.*, **20**, No. 3, 495–510.

PRUITT, D. G. and TEGER, A. I. (1967). Is there a shift toward risk in group discussion? If so, is it a group phenomenon? If so, what causes it? Paper presented at Am. Psy. Assocn, Washington, D.C.

RAIFFA, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Massachusetts: Addison-Wesley.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory.* Boston: Harvard Business School.

SAVAGE, L. J. (1954). *The Foundations of Statistics.* New York: Wiley.

—— (1961). The foundations of statistics reconsidered. In *Studies in Subjective Probability* Proc. of the Fourth (1960) Berkeley Symposium on Math. Statist. (H. E. Kyburg and H. E. Smokler, eds.), New York: Wiley.

SCHACHTER, S. (1951). Deviation, rejection and communication. *J. Ab. and Soc. Psychol.*, **46**, 190–207.

SCHACKLE, G. L. S. (1961). *Decision Order and Time in Human Affairs.* Cambridge University Press.

SCHLAIFER, R. (1969). *Analysis of Decisions under Uncertainty*. New York: McGraw-Hill.
SHEPARD, R. N. (1966). Metric structures in ordinal data. *J. Math. Psychol.*, **3**, 287–317.
SHUFORD, E. H., Jr., ALBERT, A. and MASSENGILL, H. E. (1966). Admissible probability measurement procedures. *Psychometrica*, **31**, No. 2, 125–141.
SIEGEL, S. and ZAJONC, R. B. (1967). Group risk-taking in professional decisions, *Socio.*, **30**(4), 339–349.
SILVERTHORNE, C. P. (1971). Information input and the group shift phenomenon in risk-taking. *J. Pers. & Soc. Psychol.*, **20**, No. 3, 456–461.
SLOVIC, P. (1966). Value as a determiner of subjective probability. *IEEE Trans. Hum. Fac. in Elec.*, HFE–**7**, 223–224.
SMITH, L. E. (1967). Ranking procedures and subjective probability distributions. *Man. Sci.*, **14**, 236–250.
SOUDER, W. E. (1969). The validity of subjective probabilities of success forecasts by R and D managers, *IEEE Trans. Engin. Man.* (*US*), **16**, No. 1, 35–49.
SPROWLS, R. C. (1953). Psychological–mathematical problems in relationships of lottery gambles. *Amer. J. Psychol.*, **66**, 126–130.
STAEL VON HOLSTEIN, C-A. S. (1970a). *Assessment and Evaluation of Subjective Probability Distributions*. Stockholm: The Econ. Res. Ins.
—— (1970b). A family of strictly proper scoring rules which are sensitive to distance. *J. Appl. Meteor.*, **9**, 360–364.
STONER, J. A. F. (1961). A comparison of individual and group decisions involving risk. Unpublished Masters Thesis, *M.I.T.* Sloan School of Management.
—— (1968). Risky and cautious shift in group decisions. The influence of widely held values. *J. Exper. Soc. Psychol.*, **4**, 442–459.
STRUB, M. H. (1969). Experience and prior probability in a complex decision task. *J. Appl. Psychol.*, **53**, 112–117.
TEGER, A. I. and PRUITT, D. G. (1967). Components of group risk-taking. *J. Exper. Soc. Psychol.*, **3**, 189–205.
TODA, M. (1963). Measurement of subjective probability distributions. Ins. Res. Pennsylvania: Rep. No. 3, Div. Math. Psychol.
TVERSKY, A. (1967). Additivity, utility and subjective probability. *J. Math. Psychol.*, **4**, No. 2, 175–202.
VAIL, S. (1952). Alternative calculi of subjective probabilities. In *Decision Processes* (R. M. Thrall, C. H. Coombs and R. L. Davies, eds.). New York: Wiley.
VAN NAERSSEN, R. F. (1962). A scale for the measurement of subjective probability. *Acta Psychol.*, 159–166.
VENN, J. (1888). *The Logic of Chance*, 3rd ed. London: Macmillan.
VINOKUR, A. (1971). Review and theoretical analysis of the effects of group processes upon individual and group decisions involving risk. *Psychol. Bull.*, **76**, 231–250.
WALLACH, M. A., KOGAN, N. and BEM, D. J. (1962). Group influence on individual risk-taking. *J. Ab. and Soc. Psychol.*, **65**, 75–86.
WHYTE, W. H., Jr (1956). *The Organisation Man.* New York: Simon and Shuster.
WILSON, R. B. (1968a). The theory of syndicates. *Econometrica*, **36**, No. 1, 119–131.
—— (1968b). Decision analysis in a corporation. *IEEE Trans. Sys. Sci. and Cybernetics*, SSC–**4**, No. 3, 220–226.
WINKLER, R. L. (1967a). The assessment of prior distributions in Bayesian analysis. *J. Amer. Statist. Ass.*, **62**, 776–795.
—— (1967b). The quantification of judgment: some methodological suggestions. *J. Amer. Statist. Ass.*, **62**, 1103–1120.
—— (1968). The consensus of subjective probability distributions. *Man. Sci.*, **15**, No. 2, B, 61–75.
WINKLER, R. L. and MURPHY, A. H. (1970). Non-linear utility and the probability score. *J. Appl. Meteor.*, **9**, 143–148.